New correlation coefficients for hesitant fuzzy sets

T. González-Arteaga[1] J.C.R. Alcantud[2] R. de Andrés Calle[2]

[1]Facultad de Ciencias, Universidad de Valladolid, Spain
[2]Facultad de Economía y Empresa and Multidisciplinary Institute of Enterprise (IME),
Universidad de Salamanca, Spain

Abstract

The previous correlation measures for hesitant fuzzy sets proposed in the literature only capture the strength of the correlations. We present a new approach based on the classical *Pearson correlation coefficient* for crisp values. In this way we can express not only the strength of the relationship between two hesitant fuzzy sets, but also whether they are positively or negatively associated.

Keywords: Correlation coefficient, Hesitant fuzzy sets, Scoring function, Extension Principle.

1. Introduction

The correlation coefficient is a classical tool in statistical analysis that aims to measure the joint relationship or interdependence among variables. The Pearson correlation coefficient is the most common specification and scientists and engineers often take this correlation coefficient for granted. However, practitioners often observe that precise data are not always available in many applications (see [1], [2], [3], [4], [5], [6], [7]).

Fuzzy set theory is a powerful tool to model imprecise and vague situations. Since the appearance of the fuzzy sets some extensions of this concept have been developed. Among them, we are interested in the generalization of fuzzy sets known as hesitant fuzzy sets. They were introduced by Torra [8]. Rodríguez et al. [7] is a recent authoritative survey.

The main characteristic of hesitant fuzzy sets (HFSs) is that they are defined in terms of a set of several possible membership values for each element in the domain. This makes them distinctive with fuzzy sets and other existing extensions. The HFSs allow to deal with some practical situations where establishing the membership degree is difficult and there is hesitation.

Although correlation measures for hesitant fuzzy information were discussed in Xu and Xia [9], they were not designed for HFSs. Besides, Cheng et al. [10] have defined a correlation coefficient for HFSs based on the idea of informational energy (see [11]).

In this paper, we discuss the concept of correlation for imprecise data described by HFSs. We propose new correlation coefficients based on the

concept imported from Statistics. They reflect the spirit of the classical Pearson correlation coefficient for crisp data. Consequently, they show the strength of the relationship and whether a positive or negative relationship exists between the HFSs.

The remainder of this paper is organized as follows. In Section 2, we introduce basic notation and preliminaries about hesitant fuzzy sets. Section 3 is devoted to review the idea of correlation: the fuzzy set approach and the classical point of view. Section 4 provides our proposal of new approaches for building correlation coefficients for hesitant fuzzy sets as well as their main properties. Finally, we present some concluding remarks.

2. Notation and preliminaries

We briefly describe some basic concepts related to hesitant fuzzy sets and some tools needed farther along.

The concept of HFSs was introduced as a generalization of Fuzzy sets. They permit several possible membership values in $[0, 1]$ for a single element in the reference set. The definition is cited below.

Let $\mathcal{P}^*([0, 1])$ be the set of all subsets of the unitary interval $[0, 1]$ and let $\mathcal{F}^*([0, 1])$ be the set of non-empty finite subsets of $[0, 1]$.

Definition 1 (Torra [8]) *Let $\mathcal{S}$ be a reference set. A hesitant fuzzy set (HFS) on $\mathcal{S}$ is a function $h : \mathcal{S} \to \mathcal{P}^*([0, 1])$.*

If within each image the subset in $\mathcal{P}^*([0, 1])$ is a singleton, then the HFS degenerates into a fuzzy set (with no hesitation over the membership degree), as it was mentioned in Alcantud and de Andres Calle [12].

Definition 2 (Xia and Xu [1]). *A hesitant fuzzy element (HFE) is a non-empty finite subset h of $[0, 1]$. The set of all HFEs is $\mathcal{F}^*([0, 1])$.*

A HFE represents the set of all possible membership values on an specific alternative. Based on Defintion 2, Xia and Xu [1] expressed a HFS with useful mathematical representation as follows:

$$E = \{\langle s, h_E(s)/s \in \mathcal{S}\}$$

where $h_E(s) \in \mathcal{F}^*([0, 1])$ is a HFE that expresses the possible membership degrees of s to the set E.

A notion connected with HFEs is its cardinality. Given a HFE $h_E(s)$, $l_E(s)$ stands for its cardinality, which is also known as the length of the HFE. In case of a degenerated HFS, this cardinality is 1 for all its HFEs.

We write $h_E(s) = \{(h_E(s))^{(1)}, \ldots, (h_E(s))^{(l_E(s))}\}$ where $(h_E(s))^{(j)}$ is the j−th element in $h_E(s)$ when its elements are ordered in an increasing order.

Following the recommendation in Rodríguez et at. [7], we keep the name HFS although we are concerned with the definition of typical hesitant fuzzy set in [13]. Henceforth, a HFS on $\mathcal{S}$ is a function $h : \mathcal{S} \to \mathcal{F}^*([0,1])$. When $\mathcal{S}$ is finite, HFSs are traditionally depicted in tabular representation. We use this standard description thoroughly along the paper.

Next, we recall some tools we need later to design new correlation coefficients in Section 4, that is, the Extension Principle and scoring functions.

The Extension Principle allows to export functions and operations on fuzzy sets to HFS. It was proposed by Torra and Narukawa [14] as follows:

Definition 3 (Torra and Narukawa [14]). *Let Θ be a function $\Theta : [0,1]^n \to [0,1]$ and $H = \{h_1, \ldots, h_n\}$ a set of HFSs defined on the same reference set $\mathcal{S}$. We export Θ defining $\Theta_H = \cup_{\gamma \in h_1(s) \times \ldots \times h_n(s)} \{\Theta(\gamma)\}$.*

The term score function for HFEs has been introduced by Xia and Xu [1] and then used by Farhadinia [3] with a slightly different formulation. Nevertheless the goal is the same, a way to compare HFE. For a HFE $h = \{h^1, \ldots, h^{l(h)}\}$, two established score functions are given as follows:

- Xia and Xu [1]

$$sc(h) = \frac{1}{l(h)} \sum_{j=1}^{l(h)} h^j \qquad (1)$$

- Farhadinia [3]

$$Sc(h) = \frac{\sum_{j=1}^{l(h)} \delta(j) h^j}{\sum_{j=1}^{l(h)} \delta(j)} \qquad (2)$$

where $\{\delta(1), \ldots, \delta(l(h))\}$ is a monotonic increasing sequence of index j compounded of positive values.

Therefore, $sc(h)$ in Equation (1), according to Xia and Xu [1], is the arithmetic average of the values in h and $Sc(h)$ in Equation (2), according to Farhadinia [3], is a weighted mean of the ordered values in h with weights $\dfrac{\delta(j)}{\sum_{j=1}^{l(h)} \delta(j)}$.

3. On correlation measurement

This section addresses the correlation measures from the fuzzy set approach and the classical Pearson correlation coefficient.

3.1. Correlation in Fuzzy set theory

Correlation coefficients have been a tool in fuzzy decision making models for a long time. Essentially there have been two broad approaches. One of them produces correlation coefficients into the interval $[0,1]$ and the other approach provides the corresponding correlation coefficients into the interval $[-1,1]$. So, in the former approach only the strength of relationship is evaluated, and in the latter, also a positive and negative type of a relationship is reflected.

As for the extension of the concept of correlation coefficient to the fuzzy setting, we find many papers under different perspectives. We point out some of them.

A number of authors consider an entropy-related measure from which they define correlation coefficients into the interval $[0,1]$. In this line, Dumitrescu [11] developed the so-called correlation coeffficient between two fuzzy sets and Gerstenkorn and Mańko [15] gave the correlation coefficient of two intuitionistic fuzzy sets. Their works were followed by Bustince and Burillo [16] for interval-valued intuitionistic fuzzy sets.

On other side, some authors have presented correlation coefficients between Atanassov's intuitionistic fuzzy sets ranging in the interval $[-1,1]$, like Hung and Wu (see [17]) and Szmidt et at. (see [18]) among others.

Regarding fuzzy numbers, we also find some definitions where the correlation coefficient is a value in $[0,1]$, like Yu [19] and others where the correlation coefficient ranges in $[-1,1]$, like Chiang and Lin [20]. Other definitions produce correlation coefficients that are fuzzy numbers, rather than a crisp value (e.g. see [21] and [22]).

As for hesitant fuzzy sets, a correlation coefficient has been defined in the vein of those ranging in $[0,1]$ by Chen et al. [10]. In contrast to that approach we are going to propose new correlation coefficients ranging in $[-1,1]$ in this paper.

Other extensions of the notion of correlation coefficient in fuzzy environment can be mentioned, e.g. for linguistic variables (see [23]) and for dual hesitant fuzzy sets (see [24]).

3.2. The classical Pearson correlation coefficient

Although the meaning of *correlation* seems intuitively clear, diverse attempts of formalization have appeared. We focus on the classical correlation coefficient defined by Pearson (see [25] and [26], among others).

The purpose of a correlation measure is to capture the dependence between two variables in terms of their tendency to increase or decrease either in the same or the opposite direction. If an increase in the first variable tends to come along with an increase in the other variable, then the correlation is positive. On the contrary, if a rise in the first variable tends to come along with a fall in the other one, then the correlation is negative. If there is no relation of either kind, the correlation is zero or close to zero.

The correlation coefficient between HFSs introduced in Chen et al. [10] does not fulfill the classical perspective because it does not capture the direction of the relation. However, our proposals in this paper do it.

Now we recall the definition of the Pearson correlation coefficient between two variables.

Definition 4 *Assume a sample of $n \geq 2$ subjects (or experimental units) along with a sequence of paired values from two vectors X and Y in $\mathbb{R}^n$, that is, $(x_1, y_1), \ldots, (x_n, y_n)$. The* Pearson correlation coefficient *between X and Y is given by*

$$cor(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{Y})^2}} \quad (3)$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the arithmetic means of X and Y, respectively.

3.3. Properties

The Pearson correlation coefficient has well-known properties (see e.g. [27]). To name but a few:

(1) $cor(x,y) = cor(y,x)$.
(2) $cor(x,y) \in [-1,1]$.
(3) If $x' = a + bx$ and $y' = cy + d$ where $a, b, c, d \in \mathbb{R}$, $cor(x', y') = cor(x,y)$.
(4) $cor(x,y) = 1$ or $cor(x,y) = -1$ if there exists a perfect linear relation between x and y with slope positive or negative, respectively.
(5) $cor(x,x) = 1$.

The Pearson correlation coefficient shows the strength of the association through its absolute value. And it also highlights if the relation is positive or negative by means of the sign. So, the closer the value to either $+1$ or -1, the stronger the (positive or negative) relation.

4. New correlation coefficients between two hesitant fuzzy sets

Based on the conceptual idea of the correlation described in subsection 3.2, we devote this section to develop correlation coefficients to measure the association between two HFSs.

Firstly, we propose to use the Extension Principle (Definition 3) in order to extend the correlation coefficient defined by Equation (3) to HFSs. Secondly, we provide new definitions of correlation coefficients based on the score functions (Equations (1) and (2)). We introduce both definitions as well as relevant properties.

4.1. A new correlation coefficient between two hesitant fuzzy sets based on the Extension Principle

We proceed to present the construction of a correlation coefficient between two HFS by means of the Extension Principle.

Generally speaking, the Pearson correlation coefficient is a function defined over paired real values on the n subjects, $cor : (\mathbb{R} \times \mathbb{R})^n \to [-1, 1]$. So, it is clear that paired hesitant fuzzy sets must be involved. Accordingly, we define a version of paired HFS, which is inspired in the extension principle below.

Along this subsection we let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be a reference set, and we fix two HFSs on $\mathcal{S}$, namely $X = \{\langle s, h_X(s)\rangle / s \in \mathcal{S}\}$ and $Y = \{\langle s, h_Y(s)\rangle / s \in \mathcal{S}\}$.

Definition 5 *The* **paired HFEs** *of the HFSs X and Y on $\mathcal{S}$ are the Cartesian products of $h_X(s_i) \times h_Y(s_i)$ for each s_i, $i \in \{1, \ldots, n\}$. Hence, each paired HFE comprises all pairs of elements such that the first element of the pair is from $h_X(s_i)$ and the second one is from $h_Y(s_i)$, i.e.*

$$h_X(s_i) \times h_Y(s_i) =$$
$$= \{(\gamma_X, \gamma_Y) / \gamma_X \in h_X(s_i), \gamma_Y \in h_Y(s_i)\}$$

In other words, we can also specify $h_X(s_i) \times h_Y(s_i)$ as the collection of all pairs

$$\left((h_X(s_i))^{(j)}, \ (h_Y(s_i))^{(k)} \right)$$

where $j \in \{1, \ldots, l_X(s_i)\}$ and $k \in \{1, \ldots, l_Y(s_i)\}$

In this vein, we consider the **paired HFS** *of two HFSs, X and Y, as the set of all paired HFEs for each $s_i \in \mathcal{S}$,*

$$\mathcal{R}_{HFS} = \cup_{s_i \in \mathcal{S}} h_X(s_i) \times h_Y(s_i) =$$
$$= \left\{ \left((h_X(s_i))^{(j)}, \ (h_Y(s_i))^{(k)} \right) \right\}$$

where $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, l_X(s_i)\}$ and $k \in \{1, \ldots, l_Y(s_i)\}$

The total number of pairs of values in $\mathcal{R}_{HFS}$ is

$$|\mathcal{R}_{HFS}| = \sum_{i=1}^{n}(l_X(s_i) \times l_Y(s_i))$$

Definition 6 *The* correlation coefficient *between X and Y is*

$$\rho_{HFS}(X, Y) = cor(x, y) \quad (4)$$

where (x, y) are the pairs of values in $\mathcal{R}_{HFS}$.

Hence, we state the correlation coefficient of X and Y as the value obtained when the Pearson correlation coefficient (Equation (3)) is computed from the pairs of values in $\mathcal{R}_{HFS}$. The most detailed formal expression is as follows:

$$\rho_{HFS}(X, Y) = \frac{SSC(h_X, h_Y)}{\sqrt{SS(h_X)}\sqrt{SS(h_Y)}} \quad (5)$$

where $SSC(h_X, h_Y) =$

$$= \sum_{i=1}^{n} \sum_{j=1}^{l_X(s_i)} \sum_{k=1}^{l_Y(s_i)} \left(h_X(s_i))^{(j)} - \overline{h_X} \right) \left(h_Y(s_i))^{(k)} - \overline{h_Y} \right)$$

$$\overline{h_X} = \frac{1}{|\mathcal{R}_{HFS}|} \sum_{i=1}^{n} l_Y(s_i) \left(\sum_{j=1}^{l_X(s_i)} (h_X(s_i))^{(j)} \right) \quad \text{and}$$

$$\overline{h_Y} = \frac{1}{|\mathcal{R}_{HFS}|} \sum_{i=1}^{n} l_X(s_i) \left(\sum_{j=1}^{l_Y(s_i)} (h_Y(s_i))^{(j)} \right)$$

Thus, $\overline{h_X}$ and $\overline{h_Y}$ are the arithmetic means of the corresponding values of the first and second elements of the pairs, respectively.

$$SS(h_X) = \sum_{i=1}^{n} l_Y(s_i) \left(\sum_{j=1}^{l_X(s_i)} \left((h_X(s_i))^{(j)} - \overline{h_X} \right)^2 \right)$$

$$SS(h_Y) = \sum_{i=1}^{n} l_X(s_i) \left(\sum_{j=1}^{l_Y(s_i)} \left((h_Y(s_i))^{(j)} - \overline{h_Y} \right)^2 \right)$$

Remark 1 *In the crisp case, the Pearson correlation coefficient between two variables is not defined when one of them is constant, i.e. when all its values coincide. In such a case, Equation (3) does not make sense. It is well known that the division by zero is not defined mathematically. The usual position is that under coincidences like that, it is pointless to study correlation.*

We would have an analogous situation here if one of the hesitant fuzzy sets, X or Y, has coincident membership degrees for all HFEs. For instance, $\mathcal{S} = \{s_1, s_2, s_3\}$ and $\{(s_1, \{0.7\}), (s_2, \{0.7\}), (s_3, \{0.7\})\}$. Then the corresponding $SS(h_X)$ or $SS(h_Y)$ would be 0 and the denominator in Equation (5) would be zero. Exactly as in crisp context, we assume that a such kind of coincidence does not happen.

Unlike the previous proposal in Chen et at. [10], Definition 6 allows to operate correctly without any assumption about the length of the HFEs and without adding elements to the HFEs for the purpose of the computations.

Example 1 below shows how to compute the correlation coefficient ρ_{HFS} in a practical case.
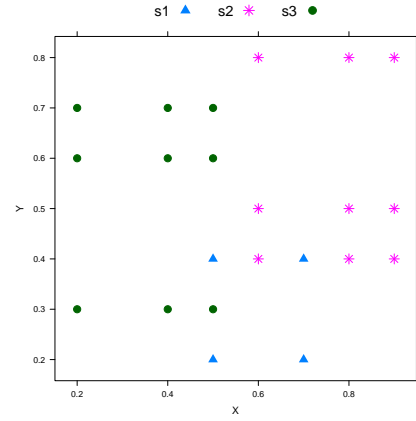

Figure 1: Scatter plot of the pairs in the Example 1

Example 1 *Consider the following example from Chen et at. [10, Ex. 2, p. 2201]. Let X and Y be two HFSs on $\mathcal{S} = \{s_1, s_2, s_3\}$ defined as*
$X = \{(s_1, \{0.7, 0.5\}), (s_2, \{0.9, 0.8, 0.6\}),$
$(s_3, \{0.5, 0.4, 0.2\})\}$ *and*
$Y = \{(s_1, \{0.4, 0.2\}), (s_2, \{0.8, 0.5, 0.4\}),$
$(s_3, \{0.7, 0.6, 0.3\})\}$.

We start building the set $\mathcal{R}_{HFS}$ of pairs (see Table 1). Figure 1 shows a scatter plot of the pairs in Table 1. Then, we use Equation (5) to calculate the correlation coefficient.
$\overline{h_X} = 0.57272$, $\overline{h_Y} = 0.5045$, $SSC(h_X, h_Y) = 0.0327$, $SS(h_X) = 1.0436$, $SS(h_Y) = 0.7695$.
Finally, $\rho_{HFS}(X, Y) = \frac{0.0327}{\sqrt{1.0436}\sqrt{0.7695}} = 0.0365$. This value is near zero in agreement with the Figure 1.

$h_X(s_1) \times h_Y(s_1)$		$h_X(s_2) \times h_Y(s_2)$		$h_X(s_3) \times h_Y(s_3)$	
0.7	0.4	0.9	0.8	0.5	0.7
0.7	0.2	0.9	0.5	0.5	0.6
0.5	0.4	0.9	0.4	0.5	0.3
0.5	0.2	0.8	0.8	0.4	0.7
		0.8	0.5	0.4	0.6
		0.8	0.4	0.4	0.3
		0.6	0.8	0.2	0.7
		0.6	0.5	0.2	0.6
		0.6	0.4	0.2	0.3

Table 1: Pairs of the Cartesian Product for the HFS of the Example 1

Note that when X and Y are degenerate HFSs, i.e., all the HFEs into X and Y have cardinality 1, and expressions (4) or (5) are applied, the correlation coefficient $\rho_{HFS}(X, Y)$ boils down to the classical Pearson correlation coefficient between two variables (Equation (3)).

In order to illustrate negative and positive associations between two HFS, we present Examples 2 and 3 below.

Example 2 *Let $\mathcal{S} = \{s_1, s_2, s_3\}$ and the HFSs on $\mathcal{S}$ given by Table 2. The corresponding paired HFEs are:*

$h_X(s_1) \times h_Y(s_1) = \{(0.05, 0.3), (0.1, 0.3), (0.15, 0.3), (0.05, 0.4), (0.1, 0.4), (0.15, 0.4), (0.05, 0.5), (0.1, 0.5), (0.15, 0.5)\}$

$h_X(s_2) \times h_Y(s_2) = \{(0.2, 0.5), (0.3, 0.5), (0.4, 0.5), (0.2, 0.6), (0.3, 0.6), (0.4, 0.6), (0.2, 0.7), (0.3, 0.7), (0.4, 0.7)\}$

$h_X(s_3) \times h_Y(s_3) = \{(0.4, 0.7), (0.5, 0.7), (0.6, 0.7), (0.4, 0.8), (0.5, 0.8), (0.6, 0.8), (0.4, 0.9), (0.5, 0.9), (0.6, 0.9)\}$

	X	Y
s_1	{0.05, 0.1, 0.15 }	{0.3, 0.4, 0.5 }
s_2	{0.2, 0.3, 0.4 }	{0.5, 0.6, 0.7 }
s_3	{0.4, 0.5, 0.6 }	{0.7, 0.8, 0.9 }

Table 2: HFSs X and Y in Example 2

Thus, we form a total number of pairs of $(3 \times 3) + (3 \times 3) + (3 \times 3) = 27$. We use them for computing $\overline{h_X} = 0.3$, $\overline{h_Y} = 0.6$, $SSC(h_X, h_Y) = 0.72$, $SS(h_X) = 0.855$ and $SS(h_Y) = 0.9$.
Now, following Definition 6, we compute the correlation coefficient through Equation (5) and we obtain $\rho_{HFS}(X, Y) = 0.8207827$.

Example 3 *In the situation of Example 2, consider now the HFSs defined in Table 3.*

	X	Y'
s_1	{0.05, 0.1, 0.15 }	{0.7, 0.8, 0.9 }
s_2	{0.2, 0.3, 0.4 }	{0.5, 0.6, 0.7 }
s_3	{0.4, 0.5, 0.6 }	{0.3, 0.4, 0.5 }

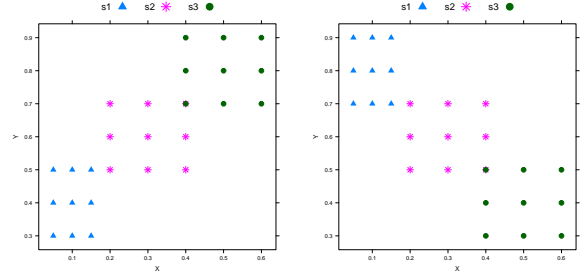Table 3: HFSs X and Y' in Example 3

The corresponding paired HFEs are:

$h_X(s_1) \times h_Y(s_1) = \{(0.05, 0.7), (0.1, 0.7), (0.15, 0.7), (0.05, 0.8), (0.1, 0.8), (0.15, 0.8), (0.05, 0.9), (0.1, 0.9), (0.15, 0.9)\}$

$h_X(s_2) \times h_Y(s_2) = \{(0.2, 0.5), (0.3, 0.5), (0.4, 0.5), (0.2, 0.6), (0.3, 0.6), (0.4, 0.6), (0.2, 0.7), (0.3, 0.7), (0.4, 0.7)\}$

$h_X(s_3) \times h_Y(s_3) = \{(0.4, 0.3), (0.5, 0.3), (0.6, 0.3), (0.4, 0.4), (0.5, 0.4), (0.6, 0.4), (0.4, 0.5), (0.5, 0.5), (0.6, 0.5)\}$

Thus, we form a total number of pairs of $(3 \times 3) + (3 \times 3) + (3 \times 3) = 27$. We use them for computing $\overline{h_X} = 0.3$, $\overline{h_Y} = 0.6$, $SSC(h_X, h_Y) = -0.72$, $SS(h_X) = 0.855$ and $SS(h_Y) = 0.9$.

(a) Positive association   (b) Negative association

Figure 2: Scatter plots of the pairs in the Example 2 (a) and in the Example 3 (b).

Now, following the Definition 6, we compute the correlation coefficient with Equation (3) from those pairs and we obtain $\rho_{HFS}(X, Y') = -0.8207827$.

Note that the strength of the relation is the same for both examples, but the sign of the correlation coefficient is different. This indicates the direction of the relationship. Figure 2 displays the scatter plot of the corresponding pairs. The direction of the association is clearly noticeable.

4.1.1. Properties

The correlation coefficient introduced in Definition 6 has the following properties. Let X and Y be two HFSs on $\mathcal{S}$.

(1) $\rho_{HFS}(X, Y) = \rho_{HFS}(Y, X)$
It states that this correlation coefficient is symmetric.
(2) $-1 \leq \rho_{HFS}(X, Y) \leq 1$
The most usual range of values for a correlation coefficient is reached with this specific definition.
(3) $\rho_{HFS}(X, Y) = 1$ or $\rho_{HFS}(X, Y) = -1$ if and only if the pairs in $\mathcal{R}_{HFS}$ are in a straight line. It shows how the extreme values of the range can be attained by this correlation coefficient.

We omit the proofs of these properties because they are straightforward from the corresponding properties of the coefficient correlation between two real numerical vectors.
Note that property (3) implies that even if the HFSs X and Y are equal, then the correlation coefficient $\rho_{HFS}(X, Y)$ does not necessarily take the value 1 as the following example shows.

Example 4 *Suppose $X = Y = \{(s_1, \{0.7, 0.5\}), (s_2, \{0.9, 0.8, 0.6\}), (s_3, \{0.5, 0.4, 0.2\})\}$. Then some algebra produces*
$\rho_{HFS}(X, Y) = \rho_{HFS}(X, X) = 0.6933798$.

If we want to design X and Y where either $\rho_{HFS}(X, Y) = 1$ or $\rho_{HFS}(X, Y) = -1$, then we need that both X and Y be degenerate HFSs (i.e., that all their HFEs have cardinality 1) and also the pairs
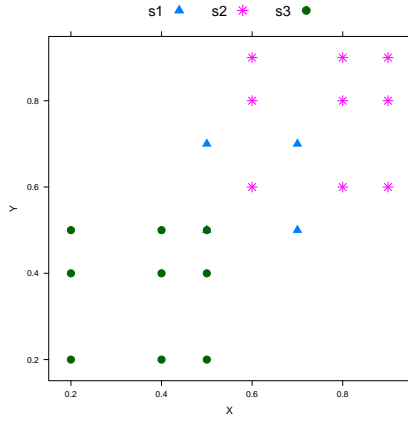
Figure 3: Scatter plot of the pairs in the Example 4

in $\mathcal{R}_{HFS}$ must be in a straight line. Otherwise, at least one paired HFE contains some elements out of an straight line because of its construction.

Note that property (5) of Pearson correlation coefficient is not verified for Definition 6. This fact is shown in Example 4.

4.2. New correlation coefficients between two hesitant fuzzy sets based on score functions

We now introduce new correlation coefficients between two hesitant fuzzy sets. It fulfills that the maximum value $(+1)$ can be reached whenever the two HFSs are equal.

In order to achieve our goal, we convert the initial problem, i.e. the quantification of the correlation between two HFSs, into a related question, namely, the computation of the correlation coefficient between the scores of the corresponding HFSs in the following precise terms.

Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be a reference set and let $X = \{\langle s, h_X(s)\rangle / s \in \mathcal{S}\}$ and $Y = \{\langle s, h_Y(s)\rangle / s \in \mathcal{S}\}$ be two HFSs on $\mathcal{S}$. We consider a score function sc, which is applied on the HFEs of X and Y. It provides $sc(h_X(s_i))$ and $sc(h_Y(s_i))$, $i \in \{1, \ldots, n\}$.

Definition 7 *The* score correlation coefficient *between the HFS, X and Y is defined as the correlation coefficient calculated from Equation (3) applied to pairs ($sc(h_X(s_i))$, $sc(h_Y(s_i))$), $i \in \{1, \ldots, n\}$. The specific expression is as follows*

$$\rho_{sc}(X,Y) = \frac{SSC(sc(h_X), sc(h_Y))}{\sqrt{SS(sc(h_X))}\sqrt{SS(sc(h_Y))}} \quad (6)$$

where

$$\overline{X}_{sc} = \frac{1}{n}\sum_{i=1}^{n} sc(h_X)(s_i) \quad \overline{Y}_{sc} = \frac{1}{n}\sum_{i=1}^{n} sc(h_Y)(s_i),$$

$$SS(sc(h_X)) = \sum_{i=1}^{n}\left(sc(h_X)(s_i) - \overline{X}_{sc}\right)^2$$

$$SS(sc(h_Y)) = \sum_{i=1}^{n}\left(sc(h_Y)(s_i) - \overline{Y}_{sc}\right)^2$$

$$SSC(sc(h_X), sc(h_Y)) =$$
$$= \sum_{i=1}^{n}(sc(h_X)(s_i) - \overline{X}_{sc})(sc(h_Y)(s_i) - \overline{Y}_{sc}).$$

Obviously, the score function plays a role in the final result. The goal of this paper is not to introduce a new definition of score function, then we use the traditional scores defined by Xia and Xu [1] and Farhadidnia [3] in order to put in practice our methodology in the following example.

Example 5 *Example 1 continued.*
Let X and Y be as in Example 1. Firstly, we use the score function given by Xia and Xu (Equation (1)) on all the HFEs in X and Y. We display the results in Table 4. We compute the correlation coefficient by means of Equation (6) and the obtained result is $\rho_{sc}(X,Y) = 0.019$.

	$sc(h_X)(s_i)$	$sc(h_Y)(s_i)$
s_1	0.6	0.3
s_2	0.7667	0.5667
s_3	0.3667	0.5333

Table 4: Results of applying the score function in Equation (1) to Example 1.

Secondly, we use the score function given by Farhadinia (Equation (2)), the associated correlation coefficient is different. The values obtained by applying that score function to the HFEs in X and Y appear in Table 5. The computations produce $\rho_{Sc}(X,Y) = 0.0607$.

$Sc(h_X)(s_i)$	$Sc(h_Y)(s_i)$
0.64	0.34
0.82667	0.64667
0.42667	0.6134

Table 5: Results of applying the score function in Equation (2) to Example 1.

4.2.1. Properties

Definition 7 satisfies the following properties. Let X and Y be two HFSs on $\mathcal{S}$ and sc be a score function.

(1) If $X = Y$, then $\rho_{sc}(X,Y) = 1$.
   It means that the correlation between two identical HFSs is always 1.
(2) $\rho_{sc}(X,Y) = \rho_{sc}(Y,X)$.
   It ensures that this score correlation coefficient is symmetric.
(3) $-1 \leq \rho_{sc}(X,Y) \leq 1$.
   It gives the range of values that the score correlation coefficient can assume.

(4) $\rho_{sc}(X, Y) = 1$ or $\rho_{sc}(X, Y) = -1$ if and only if the pairs $(sc(h_X(s_i)), sc(h_Y(s_i)))$, $i \in \{1, \ldots, n\}$ are in a straight line.

This property identifies when the extreme values are reached.

Proof 1 *Consider first property (1). If two HFSs are equal then their HFEs are equal too. Consequently, the values of the score function for the HFEs coincide as well. That is, $sc(h_X(s_i)) = sc(h_Y(s_i))$ for $i \in \{1, \ldots, n\}$ when $X = Y$. Therefore, property (1) is obvious from Equation (6).*

Properties (2) - (4) result directly from the corresponding properties of the correlation coefficient between two real numerical vectors.

We use this new score correlation coefficient between HFSs for the HFSs X and Y given in the previous Examples 2 and 3 in order to illustrate positively and negatively correlated HFSs, respectively.

Example 6 *Example 2 continued.*
Suppose we have the HFSs X and Y given in Example 2. First, we compute the score function (Equation (1)) on the HFEs in X and Y. We obtain the values in Table 6.

	s_1	s_2	s_3
$sc(h_X)(s_i)$	0.1	0.3	0.5
$sc(h_Y)(s_i)$	0.4	0.6	0.8

Table 6: Results of applying the score function (1) to Example 2.

Second, we calculate $\overline{X}_{sc} = 0.3$, $\overline{Y}_{sc} = 0.6$, $SS(sc(h_X)) = 0.08$, $SS(sc(h_Y)) = 0.08$, $SSC(sc(h_X), sc(h_Y)) = 0.08$.
Applying Equation (6) the score correlation coefficient is $\rho_{sc}(X, Y) = 1$. Therefore, this is an example where the correlation is 1 and the HFSs involved, X and Y, are not equal.

	s_1	s_2	s_3
$Sc(h_X)(s_i)$	0.12	0.34	0.54
$Sc(h_Y)(s_i)$	0.44	0.64	0.84

Table 7: Results of applying the score function 2 to Example 2.

In the same way as before, when we use the score function given by Farhadinia (Equation (2)), we obtain a different result. After using that score function, the values are in Table 7. We make the corresponding calculations $\overline{X}_{Sc} = 0.3333$, $\overline{Y}_{Sc} = 0.64$, $SS(Sc(h_X)) = 0.0883$, $SS(Sc(h_Y)) = 0.08$, $SSC(Sc(h_X), Sc(h_Y)) = 0.084$. And the score correlation coefficient is $\rho_{Sc}(X, Y) = 0.9996223$.

Example 7 *Example 3 continued.*
Now, we assume that we have the HFSs X and Y given in Example 3. Using the score function in Equation (1) we obtain the values in Table 8.

	s_1	s_2	s_3
$sc(h_X)(s_i)$	0.1	0.3	0.5
$sc(h_Y)(s_i)$	0.8	0.6	0.4

Table 8: Results of applying the score function 1 to Example 3.

After some computations we have $\overline{X}_{sc} = 0.3$, $\overline{Y}_{sc} = 0.6$, $SS(sc(h_X)) = 0.08$, $SS(sc(h_Y)) = 0.08$, $SSC(sc(h_X), sc(h_Y)) = -0.08$.
And through Equation (6), we have that the score correlation coefficient is $\rho_{Sc}(X, Y) = -1$.

	s_1	s_2	s_3
$Sc(h_X)(s_i)$	0.12	0.34	0.54
$Sc(h_Y)(s_i)$	0.84	0.64	0.44

Table 9: Results of applying the score function 2 to Example 3.

Analogously to the previous example, now we use Farhadinia's score function given by Equation (2). The computed values are in Table 9. Then, we compute again $\overline{X}_{Sc} = 0.3333$, $\overline{Y}_{Sc} = 0.64$, $SS(Sc(h_X)) = 0.0883$, $SS(Sc(h_Y)) = 0.08$, $SSC(Sc(h_X), Sc(h_Y)) = -0.084$. And the score correlation coefficient yields $\rho_{Sc}(X, Y) = -0.9996223$.

Our examples emphasize the existence of positively (see Example 6) and negatively (see Example 7) correlated HFSs.

5. Concluding remarks

In this work we introduce new correlation coefficients for hesitant fuzzy sets based on the classical definition of the Pearson correlation coefficient for crisp values. We first present the correlation coefficient that results from applying the Extension Principle directly. Then, other correlation coefficients are built through score functions. All these coefficients provide us with the magnitude of the relationship and they also distinguish whether the hesitant fuzzy sets are positively or negatively correlated. We also show their main properties.

The proposed correlation coefficients can be used in applications where the concept of correlation is required but the information is available in form of hesitant fuzzy sets, since the presence of uncertainty and ambiguity prevents classical analysis.

Acknowledgements

References

[1] M. Xia and Z. Xu. Hesitant fuzzy information aggregation in decision making. *International Journal of Approximate Reasoning*, 52(3):395 – 407, 2011.

[2] R. M. Rodríguez, L. Martínez, and F. Herrera. Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20(1):109 – 119, 2012.

[3] B. Farhadinia. A novel method of ranking hesitant fuzzy values for multiple attribute decision-making problems. *International Journal of Intelligent Systems*, 28(8):752 – 767, 2013.

[4] G. Qian, H. Wang, and X. Feng. Generalized hesitant fuzzy sets and their application in decision support system. *Knowledge-Based Systems*, 37:357 – 365, 2013.

[5] D. Yu, W. Zhang, and Y. Xu. Group decision making under hesitant fuzzy environment with application to personnel evaluation. *Knowledge-Based Systems*, 52:1 – 10, 2013.

[6] D. Yu. Triangular hesitant fuzzy set and its application to teaching quality evaluation. *Journal of Information and Computational Science*, 10(7):1925 – 1934, 2013.

[7] R. M. Rodríguez, L. Martínez, V. Torra, Z. S. Xu, and F. Herrera. Hesitant fuzzy sets: state of the art and future directions. *International Journal of Intelligent Systems*, 29(6, SI):495 – 524, 2014.

[8] V. Torra. Hesitant fuzzy sets. *International Journal of Intelligent Systems*, 25(6):529 – 539, 2010.

[9] Z. Xu and M. Xia. On distance and correlation measures of hesitant fuzzy information. *International Journal of Intelligent Systems*, 26(5):410 – 425, 2011.

[10] N. Chen, Z. Xu, and M. Xia. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. *Applied Mathematical Modelling*, 37(4):2197 – 2211, 2013.

[11] D. Dumitrescu. A definition of an informational energy in fuzzy sets theory. *Stud. Univ. Babes-Bolyai Math*, 22(2):57 – 59, 1977.

[12] J. C. R. Alcantud and R. de Andrés Calle. Hesitant fuzzy sets: the Hurwicz approach to the a nalysis of project evaluation problems. Working paper, 2014.

[13] B. Bedregal, R. Reiser, H. Bustince, C. López-Molina, and V. Torra. Aggregation functions for typical hesitant fuzzy elements and the action of automorphisms. *Information Sciences*, 255:82 – 99, 2014.

[14] V. Torra and Y. Narukawa. On hesitant fuzzy sets and decision. In *IEEE International Conference on Fuzzy Systems*, pages 1378–1382, 2009.

[15] T. Gerstenkorn and J. Mańko. Correlation of intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 44(1):39 – 43, 1991.

[16] H. Bustince and P. Burillo. Correlation of interval-valued intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 74(2):237 – 244, 1995.

[17] W. Hung and J. Wu. Correlation of intuitionistic fuzzy sets by centroid method. *Information Sciences*, 144:219 – 225, 2002.

[18] E. Szmidt, J. Kacprzyk, and P. Bujnowski. Correlation between intuitionistic fuzzy sets: Some conceptual and numerical extensions. In *IEEE International Conference on Fuzzy Systems*, pages 10–15, 2012.

[19] C. Yu. Correlation of fuzzy numbers. *Fuzzy Sets and Systems*, 55(3):303 – 307, 1993.

[20] D. Chiang and N. P. Lin. Correlation of fuzzy sets. *Fuzzy Sets and Systems*, 102(2):221 – 226, 1999.

[21] S. Liu and C. Kao. Fuzzy measures for correlation coefficient of fuzzy numbers. *Fuzzy Sets and Systems*, 128(2):267 – 275, 2002.

[22] D. H. Hong. Fuzzy measures for a correlation coefficient of fuzzy numbers under T_W (the weakest t-norm)-based fuzzy arithmetic operations. *Information Sciences*, 176(2):150 – 160, 2006.

[23] S. Ngan. Correlation coefficient of linguistic variables and its applications to quantifying relations in imprecise management data. *Engineering Applications of Artificial Intelligence*, 26(1):347 – 356, 2013.

[24] J. Ye. Correlation coefficient of dual hesitant fuzzy sets and its application to multiple attribute decision making. *Applied Mathematical Modelling*, 38(2):659 – 666, 2014.

[25] K. Pearson. Notes on the history of correlation. *Biometrika*, 13(1):25 – 45, 1920.

[26] J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59 – 66, 1998.

[27] G. W. Snedecor and W. G. Cochran. *Statistical methods, 6th Edition*. The Iowa State University Press, Ames, Iowa, 1967.